



The Effect Of Crowding On The Reading Of Program Code For Programmers With Dyslexia

McChesney, I., & Bond, RR. (2021). The Effect Of Crowding On The Reading Of Program Code For Programmers With Dyslexia. In *2021 IEEE/ACM 29th International Conference on Program Comprehension (ICPC)* IEEE.

[Link to publication record in Ulster University Research Portal](#)

Published in:

2021 IEEE/ACM 29th International Conference on Program Comprehension (ICPC)

Publication Status:

Published online: 16/03/2021

Document Version

Author Accepted version

General rights

Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact pure-support@ulster.ac.uk.

The Effect Of Crowding On The Reading Of Program Code For Programmers With Dyslexia

Ian McChesney
School of Computing
Ulster University
Newtownabbey, United Kingdom
ir.mcchesney@ulster.ac.uk

Raymond Bond
School of Computing
Ulster University
Newtownabbey, United Kingdom
rb.bond@ulster.ac.uk

Abstract—Good program layout and consistent application of style facilitates code readability and comprehension. Appropriate use of white space, in particular vertical space, is useful for organising code into logical groupings of text. Where this style is not followed then the code manifests crowding and can inhibit comprehension. When reading natural text, crowding has been recognised as disproportionately affecting the reading efficiency of dyslexic readers. We present an independent two-factorial study which examines the extent to which crowding in program code affects programmers with dyslexia. The study involved 30 participants (14 dyslexia, 16 control) reading and describing crowded and spaced versions of three Java programs. Comprehension time and accuracy were measured. An eye tracker was used to collect gaze metrics. Results are presented relating to the interaction between dyslexia and crowding. Noting the small sample size, the results show that, while there is an interaction effect on gaze metrics for some program features, the results do not suggest any significant effect whereby programmers with dyslexia are disproportionately affected by crowding in computer programs.

Index Terms—program comprehension, code layout, crowding, dyslexia, eye tracking

I. INTRODUCTION

In recent years there has been increasing discussion of the role and value of neurodiversity in the workplace, focusing on the attributes and strengths which employees with neurodiverse traits can bring to their role within a team or organisation [1], [2]. One aspect of neurodiversity is dyslexia. Individuals with dyslexia experience deficiency in word recognition, spelling, short-term memory and organisation [3]. The relationship between dyslexia and computer programming is complex. On the one hand, deficiencies such as phonological processing and recognition of novel words will likely present challenges when reading and writing code. On the other, it is suggested that people with dyslexia have enhanced skills for holistic visualisation and problem solving [4] - attributes which can be an advantage in the software development process.

Eye tracking technology offers an empirical approach to the study of dyslexia. There is a large body of work reporting such studies, e.g. Bellocchi *et al.* [5]. Similarly, eye tracking has been used to investigate how programmers read and understand program code, e.g. Sharafi *et al.* [6]. Further, a small number of studies have been undertaken which used eye tracking to

explore the gaze and program comprehension capabilities of programmers with dyslexia, e.g. McChesney and Bond [7], [8]. The study reported in this paper uses eye tracking technology to explore the effect of code layout, specifically crowding, on the gaze behaviour of programmers with dyslexia and whether this behaviour is different from that of programmers without dyslexia. Our initial findings have identified some areas of difference, however, there is no consistent pattern of behaviour to suggest that programmers with dyslexia are disproportionately disadvantaged by crowding.

The paper is organised as follows. Related work is described in Section II, focusing on the relationship between crowding and dyslexia. The experimental design is outlined in Section III, presenting the hypotheses under investigation, the experimental setup and the structure of the resulting data set. Results pertaining to the interaction of crowding and dyslexia are presented in Section IV. Section V discusses the results in relation to the hypotheses and presents some other observations from the data set. The paper concludes in Section VI with a summary of the study and opportunities for future work.

II. RELATED WORK

Informed by a review of relevant studies from educational psychology and related areas, McChesney and Bond [7] formulated hypotheses on how programmers with dyslexia would be expected to read program code. The hypotheses were based on the ways in which the dyslexic gaze typically behaves when reading natural language words and text. For example, programmers with dyslexia would be expected to have greater fixation count and greater fixation duration when reading program features compared to programmers without dyslexia. Their study speculated on the role of crowding in this regard whereby “programmers with dyslexia tended to show some inattention to those areas of the program code which, arguably, exhibited crowding – a known obstacle for dyslexic readers.”

A. Crowding and dyslexia

Dyslexia is a multi-faceted condition with a variety of subtypes and likely causes. The Dual Route Model [9] is an established model of the reading process. At the core of the

reading process is phonological processing. It is a deficiency in this process that is regarded as the key contributor to dyslexia. However, there is also a role played by ocular-motor deficiencies and it is in this context that deficiencies in the visual analysis stage can be compounded by crowding [10], [11]; for example, incorrect or impaired identification in the near foveal field and in the peripheral field. Crowding is where there are visual distractors around a target word or letter, leading to deficiency in recognising letters when surrounded by other letters [11]. Crowding can be compounded by enhanced peripheral vision in dyslexic readers, although there is uncertainty in the wider dyslexia literature as to the precise nature of such a link [4].

Reading efficiency for all readers is limited by crowding [12]. Martelli *et al.* [13] showed that, for dyslexic readers, the effect of crowding is more pronounced, requiring a disproportionately higher increase in eye fixations for word recognition. We note that, like many studies in examining developmental dyslexia, Martelli *et al.*'s study was conducted with children. The extent to which crowding in dyslexia continues in adults is not clear. For example, Doron *et al.* [14] suggest that the crowding effect amongst adults with dyslexia is minor. Furthermore, as Bellocchi *et al.* [5] showed, the effect of crowding can depend on the type of dyslexia of the reader; for example, those with moderate dyslexia were affected by crowding to a greater extent in non-word and symbol recognition than those with more severe dyslexia. However, in terms of reading, Bellocchi *et al.* [15] noted that increased inter-letter spacing can enhance reading performance in dyslexic readers and that the likely explanation for this is the attendant reduction in visual crowding. And Moll and Jones [16] showed that crowding does have a differential effect in a very specific Random Automatised Naming (RAN) task leading to an increased gaze duration.

Martelli *et al.* [13] describe how differences in typical and dyslexia gaze behaviour arising from crowding were normally apparent only within small timescales (one or two seconds) and differences often disappeared the longer the exposure time. In program comprehension tasks, such as the experiments reported in this paper, exposure times are typically 30s to 120s. As such, possible effects need to be examined not only for the overall duration of stimulus exposure but for specific time phases within the comprehension task. A temporal analysis of the comprehension task can be approached through consideration of recurrent cognitive processes of "search and deliberation" and "knowledge building" [17], also referred to as "general coherence formation processing" and "selective task-oriented" processing respectively [18]. The full modelling of such phases in program comprehension using eye tracking data is beyond the scope of this paper. Instead the approach taken is to assume "coherence formation" is associated with the initial mental processing on stimulus exposure and we will examine this at three levels of temporal granularity – the first 5s, 10s and 15s of exposure time.

We can see that the effect of crowding on the dyslexic reader is a complex phenomenon. While there is an acceptance that

crowding can inhibit reading efficiency it is not clear that it is disproportionately disadvantageous to dyslexic readers [14]. Therefore, when considering the reading of program code, it seems reasonable to keep an open mind on the interaction of crowding and dyslexia and not to presume any directional hypotheses at the outset.

B. Crowding and program code

It is widely accepted that good code layout facilitates program understanding. For example, Miara *et al.* [19] found that two-spaced indentation was optimal for program comprehension, though Bauer *et al.*'s replication study [20] questioned the precise relationship between indentation and notions of correctness and perceived difficulty. Nevertheless, program style guides are widely used in modern software engineering, e.g. [21], [22]. In particular, spacing (vertical white space) is recommended to signal logical groupings of program features.

Differences in programming language and coding style lead to different manifestations of crowding in code. As noted by McChesney and Bond [23], crowding can present in a number of ways in program source code; for example, inherent code complexity, for loop headers, long sequence constructs and poor code layout. Crowding can lead to "hidden" features. In languages such as Java, for loop header statements convey a lot of semantic information. Depending on variable naming conventions, this might be within a small lexical space. They also noted the presence of such hidden features as being a potential consequence of crowding in code [7]. As such, the study of gaze behaviour in for loops appears a worthwhile area of exploration. And this is supported by Cetin's observation [24] that the understanding of loop structures in general is a problematic aspect of program comprehension for some programmers.

When studying program comprehension using eye tracking technology, metrics of interest include fixation count, fixation duration and time to first fixation for designated areas of interest [6]. From the general dyslexia literature and the previous work in reading program code, it is not clear, in the context of code reading, if this would lead to an increase in fixation count (to process the lexical or spatial complexity arising from crowding) or a decrease in fixation count (indicative of an avoidance of complex or crowded program features) [23]. This further informs the non-directional hypotheses in the present study, positing a "difference in differences" between dyslexia and control results for the crowded and spaced programs.

III. EXPERIMENTAL DESIGN

The primary research question for our study is - does crowding disproportionately affect programmers with dyslexia in their ability to read and understand program code?

A. Hypotheses Generation

Regarding program comprehension, we are interested in comprehension time and comprehension performance. Time taken to complete the program comprehension task was measured from the first appearance of the relevant program code

on screen to when the participant said “ready” or otherwise indicated when they had finished reading and were ready to describe the program function. Both events corresponded to a keyboard event managed by the researcher.

Hypothesis 1: For a given program, the difference in **comprehension time** in the crowded program compared with the spaced program will be different for the dyslexia group compared to the control group.

After the participant had indicated they understood the purpose of the program, they were asked to describe to the researchers the function and output of the program. Graded on a 10 point scale, this was used to record the comprehension performance of the participant.

Hypothesis 2: For a given program, the difference in **comprehension performance** in the crowded program compared with the spaced program will be different for the dyslexia group compared to the control group.

In keeping with the studies reported above, we are not in a position to hypothesise on whether crowding will lead to an increase or decrease in specific gaze metrics. Since we are concerned with the visual effort required to recognise a program feature, the metrics we used are visit count and average visit duration.

Hypothesis 3: For a given program feature, the difference in **visit count** in the crowded program compared with the spaced program will be different for the dyslexia group compared to the control group.

Hypothesis 4: For a given program feature, the difference in **average visit duration** in the crowded program compared with the spaced program will be different for the dyslexia group compared to the control group.

Some of the literature [15] suggests that crowding disproportionately affects the time taken for a dyslexic reader to locate a word and, having located it, it will take longer to recognise and process. We explored recognition time by considering time to first fixation and first visit duration.

Hypothesis 5: For a given program feature, the difference in **time to first fixation** in the crowded program compared with the spaced program will be different for the dyslexia group compared to the control group.

Hypothesis 6: For a given program feature, the difference in **first visit duration** in the crowded program compared with the spaced program will be different for the dyslexia group compared to the control group.

Other gaze metrics will be used to speculatively explore the eye movements of the participants and seek out ways in which programmers with dyslexia might interact differently when reading crowded code.

B. Experimental Setup

To address the hypotheses an independent two-factorial experimental design was employed. The primary independent variables were programmer type (with dyslexia or without dyslexia) and program type (crowded or spaced). The experimental protocol and publication of eye gaze data was approved

according to the research governance policies as implemented by the Research Ethics Committee of the authors’ institution.

1) *Participants*: A total of 31 participants were recruited from the undergraduate computing programmes of the authors’ institution. Data from one participant had to be discarded due to equipment failure during a recording session. There were 14 participants with dyslexia, consisting of 3 female and 11 male. Dyslexia group mean age was 21.50 years (SD = 1.61) and mean programming experience was 3.67 years (SD = 1.87). Five participants self-reported their dyslexia as mild, 7 as moderate and 2 as severe. There were 16 participants in the control group, 2 female and 14 male. Control group mean age was 22.19 years (SD = 2.99) and mean programming experience was 4.75 years (SD = 1.96). When analysed using a two-tailed independent samples *t*-test for equality of means at a significance level of 0.05, no significant difference in age or programming experience was found between the dyslexia and control group.

2) *Programs*: During the experiment, participants were required to read and understand three small Java programs. The programs were devised so as to fit onto a single screen yet be of sufficient complexity to elicit a meaningful set of gaze data. Program 1 (5 lines of code) consists of a for loop which iterates over the integers 0 to 4, multiplying by two and displaying the result (Fig. 1 - crowded, Fig. 2 - spaced). Program 2 (13 lines of code) includes a for loop which iterates over a set of positive and negative integers, totalling the positive values and counting the zero or negative values, then displaying the result (Fig. 3 - crowded, Fig. 4 - spaced). Program 3 (17 lines of code) reverses an array of integers, displaying the array contents before and after the reverse operation (Fig. 5 - crowded, Fig. 6 - spaced). For each program, spacing was introduced between lines of code to achieve visual separation while retaining the logical coherence of the program. So, for example, in the spaced version of program 2, variable declarations remained grouped together (L01 and L02). Similarly the comment lines L03 and L04. Where the program tokens themselves introduced spacing, additional spacing was not introduced, for example L08-L09 and L10-L11.

The areas of interest (AOI) designated in the program for subsequent gaze analysis are shown in Table I. These AOIs were not shown in any way to the participants.

Each participant was presented with the three programs in order, with at least one crowded and one spaced, leading to six session types as shown in Table II. To maximise opportunity for cross-subject analysis, when session type e.g. 1 was undertaken say for a control participant, the next dyslexia participant also received session type 1. Each session started with a Welcome screen followed by eye tracking calibration. Each program was preceded by an instruction screen (e.g. “On the next screen you will see a code snippet. Review this code with a view to understanding its function. When you are satisfied that you understand the code to the best of your ability – say ready.”). When the participant indicated ready, the program was displayed again so that the participant

```

L01 int x = 5;
L02 for (int i=0; i<x; i++) {
L03     int value = i * 2;
L04     System.out.print(value + ",");
L05 }

```

Fig. 1. Program 1 - Crowded

```

L01 int x = 5;

L02 for (int i=0; i<x; i++) {

L03     int value = i * 2;

L04     System.out.print(value + ",");
L05 }

```

Fig. 2. Program 1 - Spaced

```

L01 int[] values = {5, -4, 2, -5, 1, 3};
L02 int total = 0, n = 0;

L03 // Setting up for loop
L04 // Loop counter = i
L05 for (int i = 0; i < values.length; i++) {
L06     if (values[i] > 0) {
L07         total = total + values[i];
L08     }
L09     else {
L10         n = n + 1;
L11     }
L12 }
L13 System.out.print("Total: " + total + " n: " + n);

```

Fig. 3. Program 2 - Crowded

```

L01 int[] values = {5, -4, 2, -5, 1, 3};
L02 int total = 0, n = 0;

L03 // Setting up for loop
L04 // Loop counter = i

L05 for ( int i = 0; i < values.length; i++) {

L06     if (values[i] > 0) {

L07         total = total + values[i];
L08     }
L09     else {

L10         n = n + 1;
L11     }
L12 }
L13 System.out.print("Total: " + total + " n: " + n);

```

Fig. 4. Program 2 - Spaced

TABLE I
AREAS OF INTEREST

Program	Feature	Area of Interest
1	Lines of code	L01-L05
	Identifiers	value-L03, value-L04
	for header	int i=0, i<x, i++
2	Lines of code	L01-L13
	Identifiers	n-L02, n-L10a, n-L10b, n-L13 total-L02, total-L07a, total-L07b, total-L13
	for header	int i=0, i < values.length, i++
3	Lines of code	L01-L17
	Identifiers	j-L06, j-L07, j-L11 numbers-L01, numbers-L03, numbers-L04 numbers-L06, numbers-L08, numbers-L09a numbers-L09b, numbers-L10, numbers-L15 numbers-L16 temp-L08, temp-L10
	for header	int i=0-L03, i < numbers.length-L03, i++-L03 int i=0-L07, i < j-L07, i++-L07 int i=0-L15, i < numbers.length-L15, i++-L15

TABLE II
EXPERIMENT SESSION TYPES
(C)ROWDED, (S)PACED

Session Type	Program 1	Program 2	Program 3
1	C	C	S
2	C	S	C
3	S	C	C
4	S	S	C
5	S	C	S
6	C	S	S

could verbally describe its function, with audio recorded. Gaze metrics were collected for both instances of the program on display. In this paper, we are reporting data pertaining to the first instance, i.e. the comprehension task rather than the explanation task. On completion of the explanation, an evaluation prompt screen was displayed with the text “How confident are you in your understanding of this code? 1 = Low Confidence, 10 = High Confidence”. For each participant a comprehension grading of the program was also completed by the researchers (on a scale 1-10).

3) *Instrumentation*: Experiment sessions were undertaken in a room in which only a researcher and participant were present. Each session lasted approximately 20 minutes. The prompts and programs described above were presented on a Hanns-G HL249DPB 23.6” 1920 x 1080 monitor. Eye tracking was undertaken using a Tobii X3-120. Participants were seated at a gaze distance of approximately 65cm. Text in programs 1 and 2 was displayed at a font size of 27/72 inches (9.53mm, 0.85 degrees) and in program 3 at a font size of 18/72 inches (6.35mm, 0.53 degrees). The specification accuracy of the Tobii X3-120 is 0.40 degrees. The calibration accuracy averaged across participants was 0.58 degrees. Tobii Pro Lab v1.145 was used to analyse the recording data.

```

L01 int[] numbers = {1, 2, 3, 4, 5};

L02 System.out.println("-----");
L03 for (int i = 0; i < numbers.length; i++) {
L04     System.out.print(numbers[i] + " ");
L05 }

L06 int j = numbers.length - 1;
L07 for (int i = 0; i < j; i++) {
L08     int temp = numbers[i];
L09     numbers[i] = numbers[j];
L10     numbers[j] = temp;
L11     j--;
L12 }

L13 System.out.println();
L14 System.out.println("-----");
L15 for (int i = 0; i < numbers.length; i++) {
L16     System.out.print(numbers[i] + " ");
L17 }

```

Fig. 5. Program 3 - Crowded

```

L01 int[] numbers = {1, 2, 3, 4, 5};

L02 System.out.println("-----");

L03 for (int i = 0; i < numbers.length; i++) {

L04     System.out.print(numbers[i] + " ");
L05 }

L06 int j = numbers.length - 1;

L07 for (int i = 0; i < j; i++) {

L08     int temp = numbers[i];
L09     numbers[i] = numbers[j];
L10     numbers[j] = temp;
L11     j--;
L12 }

L13 System.out.println();
L14 System.out.println("-----");

L15 for (int i = 0; i < numbers.length; i++) {

L16     System.out.print(numbers[i] + " ");
L17 }

```

Fig. 6. Program 3 - Spaced

C. The Data Set

The data set from the experiment is available on figshare¹ and can be used for further analysis or replication. Items available at the repository include:

- 1) Anonymised participant data such as participant age, programming experience, dyslexia status, fatigue and total session time.

¹<https://doi.org/10.6084/m9.figshare.c.5304473.v1>

TABLE III
NUMBER OF PARTICIPANTS PER FACTOR

Program	Group	Program Type	
		Crowded	Spaced
1	Dyslexia	8	6
	Control	8	8
2	Dyslexia	7	7
	Control	9	7
3	Dyslexia	7	7
	Control	8	8

- 2) Stimuli prompt screens and programs as image files.
- 3) Tobii raw data export files for each participant (30 files in total). Files are identified by pseudo-randomised participant ID. The format of the output file is given in the Tobii X3-Pro User Manual² and includes data such as calibration and gaze point coordinates for each fixation and saccade. The data files also record the area of interest where the gaze event occurred.
- 4) For each program a file is provided containing the eye gaze metrics calculated by Tobii Pro Lab (three files in total). The full set of metrics is defined in the Tobii X3-Pro User Manual and includes metrics for each AOI such as average fixation duration, number of fixations, time to first fixation and number of visits. Additional files of identical format are provided containing the eye metrics as generated for the first 5s, 10s and 15s of each session (three additional files per program).

Dyslexia and control groups were presented with programs according to the breakdown shown in Table III. Where opportunity for further recruitment arises, the data set will be extended with additional participants.

IV. RESULTS - THE INTERACTION OF CROWDING AND DYSLEXIA

To answer our hypotheses, data was analysed in SPSS Version 26 using two-way independent ANOVA ($\alpha = 0.05$) for independent variables programmer type (dyslexia, control) and program type (crowded, spaced). In this paper we limit our analysis and discussion to the interaction between the two independent variables. Where results are significant ($p \leq .050$) we report the F -statistic, significance level p and effect size η_p^2 .

A. Comprehension

Hypothesis 1: Comprehension time. The mean time spent by participants reviewing each program is shown in Table IV, showing no significant interaction between dyslexia and crowding on comprehension time.

Hypothesis 2: Comprehension performance. Based on the participant's verbal description of the purpose and output of the program, comprehension was subjectively assessed by the researchers on a 10 point scale. The criteria for grading were: 10,9 = correct description; 8,7 = one minor mistake reflecting a

²<https://www.tobiiipro.com>

TABLE IV
MEAN COMPREHENSION TIME AND SD (SECONDS)
(D)YSLEXIA, (C)ONTROL

Prog.	Group	Program Type		$F(1,26)$	p
		Crowded	Spaced		
1	D	36.24 (24.90)	39.93 (35.87)	0.13	.725
	C	34.02 (15.93)	31.70 (12.88)		
2	D	96.21 (76.28)	104.21 (69.52)	0.09	.769
	C	65.66 (25.56)	62.16 (24.02)		
3	D	118.68 (70.85)	179.09 (94.17)	0.49	.489
	C	105.39 (53.71)	129.12 (64.54)		

TABLE V
MEAN COMPREHENSION SCORE AND SD (1-10)
(D)YSLEXIA, (C)ONTROL

Program	Group	Program Type		$F(1,26)$	p
		Crowded	Spaced		
1	D	7.00 (2.27)	5.83 (2.79)	0.00	.982
	C	7.50 (2.50)	6.38 (2.62)		
2	D	8.57 (2.07)	6.14 (1.68)	2.82	.105
	C	6.78 (3.03)	7.29 (2.29)		
3	D	5.14 (3.02)	6.71 (2.36)	0.49	.492
	C	4.25 (3.58)	7.38 (3.02)		

small gap in comprehension; 6,5 = output / function described incompletely but evidence of overall understanding; 4,3 = incorrect output and limited understanding; 2 = completely wrong understanding of program; 1 = no answer. Table V shows the results, again showing no significant interaction between dyslexia and crowding on comprehension performance.

B. Visit Count

Hypothesis 3: Visit count. For this metric our hypothesis did not predict a directional difference. For programs 1, 2 and 3 there was no significant interaction between dyslexia and crowding for any of the areas of interest - lines of code, identifiers or for statement headers.

When limiting the time of interest to the first 5s, 10s and 15s of the program displayed on screen, we observed significant differences for AOIs in programs 1 and 2 as shown in Table VI.

C. Visit Duration

Hypothesis 4: Average visit duration. For program 1 there was a no significant interaction effect. For program 2 we have significant values as shown in Table VII. For program 3 there was no significant interaction effect.

When limiting time of interest to first 5s, 10s and 15s phases, for average visit duration we have significant interactions for programs 1 and 2 as shown in Table VIII.

D. Time to First Fixation

Hypothesis 5: Time to first fixation³. Here we are interested in how long it takes to find or locate program features. for program 1 there was no interaction effect relating to the for statement header (L02) nor any AOI. For program 2 there

³Tobii metric: time to first whole fixation

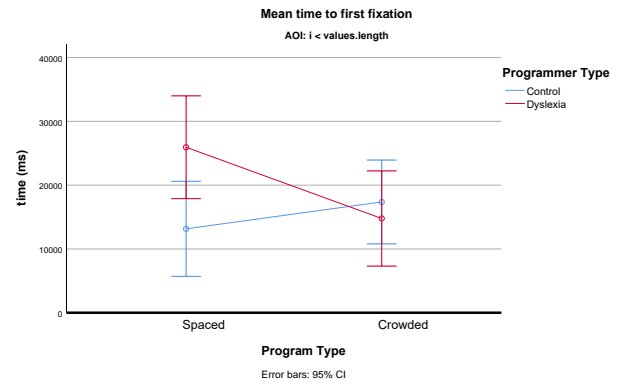


Fig. 7. Program 2: Interaction effect on time to first fixation for $i < \text{values.length}$

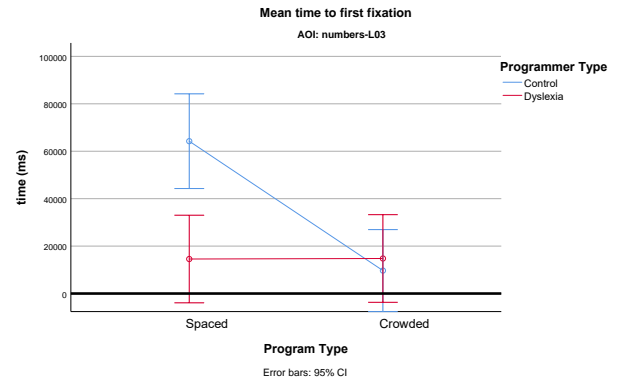


Fig. 8. Program 3: Interaction effect on time to first fixation for numbers-L03

was no interaction effect relating to the for statement header (L05). However, there was an interaction effect on the $i < \text{values.length}$ AOI within the for header, with the difference in differences significantly smaller for the crowded program (see Fig. 7). There were other program 2 AOIs where this was also true - L03, L04, L06, and total-L02. That is, in program 2 we observed a consistent pattern of time to first fixation for the dyslexia group shorter in duration than the control group in the crowded program (see Table IX). For program 3, we observed no significant interaction effect on the for statement headers L03, L07, L15. However, for numbers-L03, we do have an interaction effect, with the dyslexia group showing similar times across the two program types but the control group being significant lower in the crowded program. (Fig. 8).

E. First Visit Duration

Hypothesis 6: First visit duration. There were no significant interaction effects for program 1. For program 2, there was no significant interaction effect observed relating to the for loop (L05). We do see an interaction effect for total-L02. For program 3, there is no interaction effect relating to any of the three for loops, while for L10, there is a significant effect (see Table X).

TABLE VI
MEAN VISIT COUNT IN FIRST 5s, 10s, 15s
(SD IN BRACKETS)

Program	Time Phase	AOI	$F(1,26)$	p	η_p^2	Dyslexia		Control	
						Crowded	Spaced	Crowded	Spaced
1	5s	L03	6.10	.020	0.19	0.50 (0.76)	0.67 (0.52)	1.63 (1.30)	0.25 (0.46)
		value-L03	4.27	.049	0.14	0.50 (0.76)	0.67 (0.52)	0.88 (0.99)	0.00 (0.00)
2	5s	L04	4.85	.037	0.16	1.00 (1.00)	0.00 (0.00)	0.22 (0.44)	0.14 (0.38)
		L06	6.44	.017	0.20	1.00 (1.16)	0.00 (0.00)	0.11 (0.33)	0.29 (0.49)
	10s	values-L06	5.73	.024	0.18	1.43 (1.51)	0.00 (0.00)	0.11 (0.33)	0.00 (0.00)
		L06	8.31	.008	0.24	1.43 (1.40)	0.14 (0.38)	0.44 (1.01)	1.29 (0.95)
		values-L06	4.23	.050	0.14	2.00 (2.31)	0.14 (0.38)	0.44 (1.01)	0.57 (0.79)

TABLE VII
MEAN AVERAGE VISIT DURATION (MS) - PROGRAM 2
(SD IN BRACKETS)

AOI	$F(1,26)$	p	η_p^2	Dyslexia		Control	
				Crowded	Spaced	Crowded	Spaced
L02	4.89	.036	0.16	506.71 (195.18)	507.43 (256.77)	377.44 (169.06)	708.00 (191.33)
L13	5.12	.032	0.17	624.86 (552.85)	1958.00 (1643.26)	738.33 (735.90)	501.57 (369.31)
n-L02	6.54	.017	0.20	280.86 (66.01)	155.55 (135.91)	158.78 (103.18)	228.00 (98.50)
total-L02	4.96	.035	0.16	168.86 (106.34)	189.86 (58.21)	160.78 (38.40)	318.43 (116.49)

TABLE VIII
MEAN AVERAGE VISIT DURATION (MS) IN FIRST 5s, 10s, 15s
(SD IN BRACKETS)

Program	Time Phase	AOI	$F(1,26)$	p	η_p^2	Dyslexia		Control	
						Crowded	Spaced	Crowded	Spaced
1	5s	value-L03	4.51	.043	0.15	74.00 (136.27)	136.00 (139.21)	131.25 (156.30)	0.00 (0.00)
2	5s	L02	6.94	.014	0.21	429.14 (475.78)	99.29 (107.60)	241.22 (284.16)	516.29 (279.33)
		total-L02	7.41	.011	0.22	223.86 (376.32)	59.00 (74.54)	96.11 (97.85)	339.00 (142.54)
		L04	5.03	.034	0.16	191.57 (214.92)	0.00 (0.00)	28.78 (60.10)	21.43 (56.70)
	10s	L03	4.33	.048	0.14	525.00 (425.48)	224.71 (257.02)	176.89 (368.83)	660.57 (843.28)
		total-L02	6.77	.015	0.21	195.29 (223.19)	117.29 (86.98)	160.33 (74.77)	366.29 (175.47)
		values-L06	6.43	.018	0.20	109.57 (116.48)	0.00 (0.00)	4.44 (13.33)	0.00 (0.00)
	15s	L06	7.14	.013	0.22	542.29 (683.41)	85.71 (226.78)	63.89 (177.16)	458.86 (505.51)
		total-L02	9.19	.005	0.26	222.00 (205.80)	112.29 (79.62)	170.11 (53.35)	367.86 (173.45)

TABLE IX
MEAN TIME TO FIRST FIXATION (MS) - PROGRAM 2 AND PROGRAM 3
(SD IN BRACKETS)

Program	AOI	$F(1,26)$	p	η_p^2	Dyslexia		Control	
					Crowded	Spaced	Crowded	Spaced
2	i < values.length	4.58	.042	0.16	14775.57 (7993.13)	25942.17 (16554.40)	17369.33 (7351.21)	13156.57 (4217.59)
	L03	5.06	.033	0.16	5228.86 (3124.13)	10963.14 (10892.80)	11888.22 (4426.51)	7531.00 (2655.23)
	L04	11.14	.003	0.30	4889.71 (3819.04)	23627.29 (16603.57)	11441.11 (6399.30)	8038.00 (2923.69)
	L06	6.17	.016	0.20	13308.86 (11843.14)	29735.86 (23073.43)	21585.67 (4843.74)	13073.29 (7137.46)
	total-L02	4.74	.039	0.16	3351.33 (4425.78)	11204.86 (12663.05)	5590.67 (4241.07)	2028.00 (1618.03)
3	numbers-L03	9.26	.006	0.28	14773.29 (10763.47)	14550.71 (13851.54)	9673.50 (5283.82)	64236.83 (47805.56)

TABLE X
MEAN FIRST VISIT DURATION (MS) - PROGRAM 2 AND PROGRAM 3
(SD IN BRACKETS)

Program	AOI	$F(1,26)$	p	η_p^2	Dyslexia		Control	
					Crowded	Spaced	Crowded	Spaced
2	total-L02	6.45	.018	0.21	300.00 (370.99)	153.43 (39.66)	154.89 (102.89)	383.29 (175.38)
3	L10	6.68	.016	0.21	404.17 (409.95)	175.71 (190.94)	160.38 (58.63)	466.63 (344.45)

V. DISCUSSION

A. Comprehension time and performance

Considering hypotheses 1 and 2 concerning comprehension time and comprehension performance, our results show no significant interaction between dyslexia and crowding relating to these two dependent variables. However, regarding comprehension time across the two program types, the dyslexia group did take longer to complete program 2 (dyslexia mean = 100.21s, SD = 70.24; control mean = 64.13s, SD = 24.13; $F(1,26) = 3.50$, $p = .073$), trending towards a main effect of dyslexia on the time to read the code. We can understand how program 1 is not associated with any effect on reading time as it is small and less complex than program 2. However, while program 2 took longer to comprehend for the dyslexia group, it is not apparent why program 3 does not also lead to a longer comprehension time for the dyslexia group given it is the longest of the three programs. There is a degree of code pattern repetition in program 3 and this could ameliorate the effect of program size. A further difference is that program 2 is the only to contain “natural text” in the form of comments – though lines L03 and L04 did not show any significantly longer visit duration for the dyslexia group.

Regarding comprehension performance across the two programmer types, the crowded version of program 3 returned a significantly lower comprehension score (crowded mean = 4.67, SD = 3.24; spaced mean = 7.07, SD = 2.66; $F(1,26) = 4.43$, $p = .045$), suggesting a main effect being that the crowded program variant is more difficult to comprehend. Why this was not evident for programs 1 and 2 could be related to their relative size, suggesting that if crowding is going to be a factor in comprehension, it is more likely to exhibit the larger the program.

B. Gaze metrics

Hypotheses 3 and 4 are concerned with ways in which the gaze behaviour of programmers with dyslexia would be different across the two program types. Overall, such behaviour was not apparent. On visit count (hypothesis 3) there was no significant interaction effect in any of the three programs. On average visit duration (hypothesis 4), there was no significant interaction effect for program 1 or 3. In program 2, L02 (integer variable declarations) did exhibit an interaction effect whereby the control group spent significantly longer on this line in the spaced program (Fig. 9). And yet the last line in the program L13 (output statement) exhibited average visit duration in the opposite direction with the spaced program attracting more dyslexia gaze (Fig. 10).

The sparse and conflicting nature of these results suggest that crowding does not lead to a significantly different attention gaze for the dyslexia group compared to the control group. According to Martelli *et al.* [13], when stimuli were presented briefly, large differences in contrast thresholds for effective letter identification were observed but the threshold levels became similar with longer exposure time. Though their study concerned single letter and single word identification, in contrast to our stimulus of an entire program, exposure time could

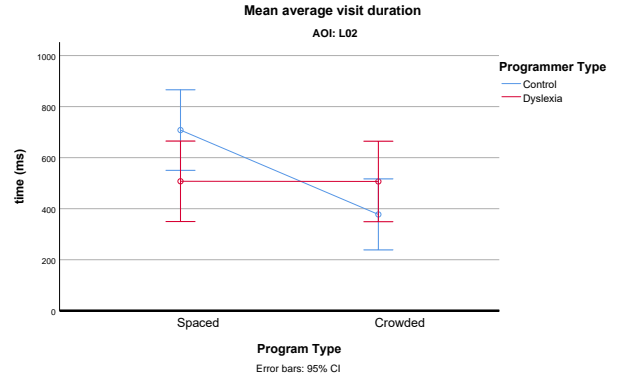


Fig. 9. Program 2: Interaction effect on visit duration for L02

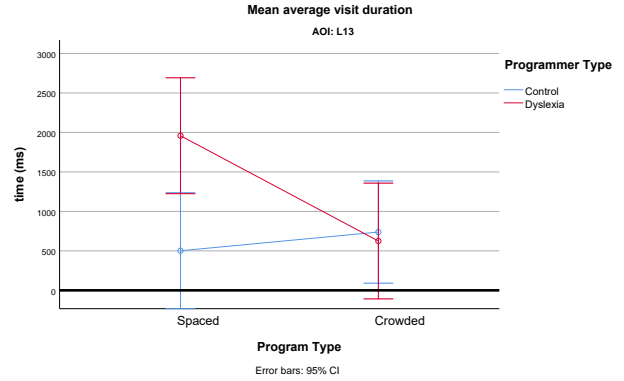


Fig. 10. Program 2: Interaction effect on visit duration for L13

explain why, by the end of each program comprehension task, there was no significant difference in the gaze patterns. Using Martelli *et al.*'s phrase – differences between dyslexia and control participants were “washed out” over longer exposure times.

In view of this, metrics were also analysed at 5s, 10s and 15s intervals from stimulus onset to see if differences were apparent at these early search stages of the comprehension task. Looking at program 1, the value-L03 AOI drew out a significant interaction effect in visit count and average visit duration in the first 5s, with the dyslexia group showing a significantly lower number of visits compared to the control group in the crowded program (Fig. 11). We note that this identifier is visually in the middle of the code space. However, beyond this early search phase such differences in gaze were no longer apparent.

For program 2 we see more complex behaviour. The total-L02 identifier shows significantly higher average visit duration for the dyslexia group in the crowded program over the first 5s, 10s and 15s of the stimuli. Something about this identifier is drawing the dyslexic gaze in the search phase. The visit count is not significantly higher, so it is the gaze duration (attention) which is different. The two comment lines in the program (L03 and L04) also show an interaction effect, the

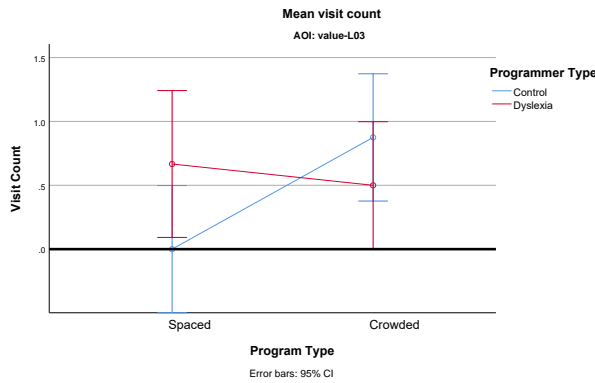


Fig. 11. Program 1: Interaction effect on visit count (first 5s) for values-L03

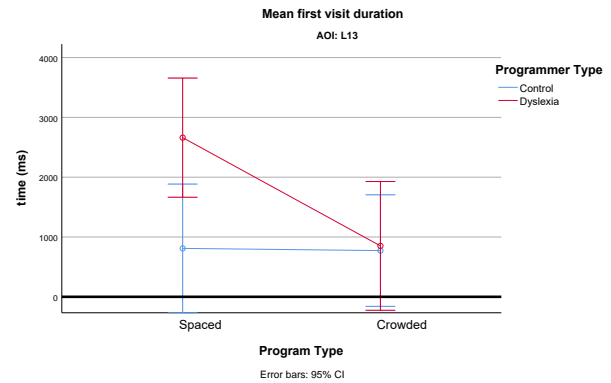


Fig. 12. Program 2: Interaction effect on first visit duration for L13

dyslexia group with notably higher average visit duration in the crowded program in the first 5s (L04 significant with $p = .034$ and L03 trending at $p = .057$). Visit count also has a significant difference for L04 in the 5s interval. This behaviour extends into the 10s phase for L03 only. The if statement (L06) similarly shows significantly higher dyslexia visit count in the 10s and 15s phases, with the then clause (L07) trending ($p = .056$) within the 15s phase. So as a program feature, the if statement is leading to more visits for the dyslexia group in the crowded program – and not the for header (L05) which had been speculated by McChesney and Bond [23]. We note again, however, that L06 is a feature which is visually at the centre of the code space. It is obviously an important part of the program’s logic – but nonetheless there is this interesting visual aspect to its location and purpose which might be important.

Interestingly, for program 3 there were no significant differences to be observed even in the search phases of the comprehension processes. Why is this so? Maybe the larger size of the program dispersed any such differences even in the first 15s? The larger the program space the less chance of any one AOI drawing significant gaze behaviour?

Hypotheses 5 and 6 explored if any program features would see the programmer with dyslexia take longer to locate or “recognise” a program feature and, having done so, if their first visit would be longer in the crowded program to deal with excessive crowding. For program 1, time to first fixation showed no interaction effect on the for header or any AOI, suggesting no crowding effect for dyslexia on program feature recognition.

In program 2, time to first fixation did show a significant interaction effect in one aspect of the for header, namely the loop condition ($i < \text{values.length}$), with the dyslexia time significantly lower than the control group in the crowded program. First visit duration did not show any interaction effect on the for header. Time to first fixation was also lower for a number of other program features – the comment lines (L03, L04) and the if header (L06), with the dyslexia group finding these features more quickly than the control group in the crowded program. In program 2 there were no AOIs which

required the dyslexia group more time to find compared to the control. For total-L02 there was an interaction effect whereby the dyslexia group had significantly greater first visit duration in the crowded program. We note a trending interaction relating to first visit duration for the last line of program 2 (L13), where we see the dyslexia group having greater first visit duration for L13 in the spaced program (dyslexia mean = 2661.71, SD = 1950.39; control mean = 809.67, SD = 830.29) and dyslexia and control groups with approximately the same values in the crowded program (dyslexia mean = 851.33, SD = 1185.29; control mean = 772.88, SD = 761.66; $F(1,26) = 3.22$, $p = .086$). As the last line in the program, the spaced version sees the dyslexia group spend significantly longer on this AOI (Fig. 12).

In program 3 there are three for loops. None of the for loop header lines were associated with any significant difference on time to first fixation nor first visit duration – suggesting again the for loop per se is not a feature associated with any interaction effect on gaze behaviour. However, considering the first for loop (L03), the numbers identifier in the loop condition (numbers-L03) did exhibit an interaction on time to first fixation, suggesting some delay for the dyslexia group in accessing the information associated with the for loop. First visit duration on L10 (penultimate statement of the central for loop) shows an interaction effect, with the dyslexia group spending more time on this AOI in the crowded program. It is reasonable to describe L10 as located in a crowded area of the program.

In summary, while there are some interaction effects to be observed within specific AOIs across all three programs, overall these do not suggest a significant difference in differences between the factors, in turn suggesting that crowding does not lead to any detrimental performance in visual access to, or recognition of, program features for the dyslexia group.

C. Other Observations - Exploratory metrics

In addition to the metrics relating to our hypotheses, other eye tracking measurements are available which might point to differences in gaze behaviour. When we explored the full range of fixation and visit metrics available for program 1,

TABLE XI
SIGNIFICANT EXPLORATORY METRICS

Program	Gaze Metric*	AOI	$F(1,26)$	p	η_p^2	Dyslexia		Control	
						Crowded	Spaced	Crowded	Spaced
2	min. VD	i++	5.91	.024	0.22	103.40 (32.55)	200.00 (104.23)	213.14 (81.20)	144.50 (89.29)
	max. VD	n-L13	9.54	.013	0.52	286.00 (29.61)	140.00 (53.75)	96.00 (21.92)	117.00 (.)
	min. FD	i < values.length	5.37	.029	0.18	76.43 (16.10)	100.00 (35.17)	79.78 (11.08)	70.43 (6.29)
3	min. FD	numbers-L06	8.58	.007	0.26	115.00 (39.17)	70.57 (9.45)	83.43 (11.79)	103.12 (39.43)
	max. VD	numbers-L16	4.98	.044	0.28	612.50 (426.96)	83.33 (8.51)	193.75 (81.23)	158.33 (118.02)

*VD = visit duration, FD = fixation duration

analysing again at the $p \leq 0.05$ significance level, there was no significant interaction between dyslexia and crowding for any such metrics on the AOIs.

Analysis for program 2 did yield some significant results (Table XI). We contrast the interaction effect on minimum visit duration on i++, where the dyslexia group was lower in the crowded program, with maximum visit duration on n-L13, where dyslexia value was greater in the crowded program. So whenever the dyslexia group "glance" (minimum values for visit duration) at this AOI, they do so for a significantly shorter time in the crowded program compared to the spaced program. Whereas with the final feature in the program (n-L13) the dyslexia group dwell on this significantly longer than the control group.

Staying with the for loop in program 2, we also notice that glances on another element of the header (i < values.length) there is an interaction effect on minimum fixation duration, with dyslexia values lower in the crowded program. When the programmer with dyslexia glances at the for header, again the glances are noticeably of shorter duration in the crowded program.

For program 3, we can observe that some gaze behaviour related to the numbers identifier shows a significant interaction effect. For numbers-L03 in the first for loop header, we have already noted the significant difference in differences on time to first fixation. For numbers-L06, we have an interaction effect on minimum fixation duration, with a higher value for the dyslexia group in the crowded program. In the final occurrence of the numbers identifier in program 3 (numbers-L16), we notice the maximum visit duration was significantly greater for dyslexia group in the crowded program, and this is consistent with observations relating to the last line of program 2. So we have, in programs 2 and 3, a tendency for the dyslexia group to dwell on the end of the programs more so than the control group but only in the crowded version of the program. Are the "edges" of the code space drawing the gaze of the programmer with dyslexia more so than control group? And if so, what effect if any is this having on comprehension?

D. Study Limitations

The sample size in the study is small leading to low statistical power for many of the results. The study group (programmers with dyslexia) is by nature a small group from which to recruit participants. Estimates for the prevalence of

dyslexia within the community are in the range 3%-17% with most accepted figures being less than 10% [25]. A value of less than 10% is consistent with the percentage within the target recruitment group for this study. With the 14 dyslexia participants we managed to recruit, the two-factorial design meant that program types were being analysed with between 6-8 participants. Straightaway this limits the statistical power of results and increases the likelihood of not detecting interactions effects which are actually present in the population. Nevertheless, the statistically significant results described and their effect sizes stand on their own as pointers to areas for further exploration in this area.

We have limited our analysis in this paper to the interaction between dyslexia and crowding. Where there is no interaction effect to be observed, we may proceed to study the main effect of dyslexia or crowding. Our consideration of comprehension time and comprehension performance in Section V-A above illustrates this approach. As such, there is further worthwhile analysis to be conducted in this regard; for example, exploring differences in gaze behaviour between the spaced and crowded programs for all participants, and wider analysis of gaze differences between programmers with and without dyslexia.

VI. CONCLUSION

Though much of the dyslexia literature suggests that crowding is disproportionately detrimental to dyslexic readers, there is debate concerning the precise mechanisms at work, including the extent to which this is true for adults with dyslexia. In considering the effect of crowding in program code on the programmer with dyslexia, our analysis answers our research question by suggesting that overall, the programmer with dyslexia shows no disproportionate deficiency. And while our results on this data set are preliminary, nevertheless our work here reinforces earlier findings [7] that dyslexia is not a detrimental condition when it comes to programming, at least in terms of code reading and program comprehension.

There are still research questions to be answered that are not within the scope of this paper. For example, to what extent does crowding and dyslexia affect non-linear reading patterns in program code? Is there a temporal pattern to be observed in how programmers with dyslexia search the information space in a crowded versus spaced program? When verbally explaining a program's purpose, does the programmer with dyslexia exhibit any difference in gaze behaviour or

explanation time? Is there a link between Gestalt principles of program code layout (including crowding) and eye gaze behaviour of programmers with and without dyslexia? We hope these questions will form the basis for further work on the data set.

ACKNOWLEDGEMENTS

We are grateful to Dr Victoria Simms for conversations on the study design and ethics approval. And to the participants who consented to their gaze data being used for this and further study.

REFERENCES

- [1] A. Krzeminska, R. D. Austin, S. M. Bruyère, and D. Hedley, "The advantages and challenges of neurodiversity employment in organizations," *Journal of Management & Organization*, vol. 25, no. 4, pp. 453–463, 2019.
- [2] S. Wille and D. Sajous-Brady, "The inclusive and accessible workplace," *Communications of the ACM*, vol. 61, no. 2, pp. 24–26, 2018.
- [3] S. Wallace, *A dictionary of education*. OUP Oxford, 2015.
- [4] J. W. Gilger, K. Allen, and A. Castillo, "Reading disability and enhanced dynamic spatial reasoning: A review of the literature," *Brain and Cognition*, vol. 105, pp. 55–65, 2016.
- [5] S. Bellocchi, M. Muneaux, M. Bastien-Toniazzo, and S. Ducrot, "I can read it in your eyes: What eye movements tell us about visuo-attentional processes in developmental dyslexia," *Research in developmental disabilities*, vol. 34, no. 1, pp. 452–460, 2013.
- [6] Z. Sharafi, B. Sharif, Y.-G. Guéhenec, A. Begel, R. Bednarik, and M. Crosby, "A practical guide on conducting eye tracking studies in software engineering," *Empirical Software Engineering*, vol. 25, no. 5, pp. 3128–3174, 2020.
- [7] I. McChesney and R. Bond, "Eye tracking analysis of computer program comprehension in programmers with dyslexia," *Empirical Software Engineering*, vol. 24, no. 3, pp. 1109–1154, 2019.
- [8] —, "Observations on the linear order of program code reading patterns in programmers with dyslexia," in *Proceedings of the Evaluation and Assessment in Software Engineering*, 2020, pp. 81–89.
- [9] M. Coltheart, "Modeling reading: The dual-route approach," *The science of reading: A handbook*, pp. 6–23, 2005.
- [10] E. Moores, R. Cassim, and J. B. Talcott, "Adults with dyslexia exhibit large effects of crowding, increased dependence on cues, and detrimental effects of distractors in visual search tasks," *Neuropsychologia*, vol. 49, no. 14, pp. 3881–3890, 2011.
- [11] S. Gori and A. Facoetti, "How the visual aspects can be crucial in reading acquisition: The intriguing case of crowding and developmental dyslexia," *Journal of Vision*, vol. 15, no. 1, pp. 8–8, 2015.
- [12] D. G. Pelli, K. A. Tillman, J. Freeman, M. Su, T. D. Berger, and N. J. Majaj, "Crowding and eccentricity determine reading rate," *Journal of Vision*, vol. 7, no. 2, pp. 20–20, 2007.
- [13] M. Martelli, G. Di Filippo, D. Spinelli, and P. Zoccolotti, "Crowding, reading, and developmental dyslexia," *Journal of Vision*, vol. 9, no. 4, pp. 14–14, 2009.
- [14] A. Doron, M. Manassi, M. H. Herzog, and M. Ahissar, "Intact crowding and temporal masking in dyslexia," *Journal of Vision*, vol. 15, no. 14, pp. 13–13, 2015.
- [15] S. Bellocchi, D. Massendari, J. Grainger, and S. Ducrot, "Effects of inter-character spacing on saccade programming in beginning readers and dyslexics," *Child Neuropsychology*, vol. 25, no. 4, pp. 482–506, 2019.
- [16] K. Moll and M. Jones, "Naming fluency in dyslexic and nondyslexic readers: Differential effects of visual crowding in foveal, parafoveal, and peripheral vision," *Quarterly Journal of Experimental Psychology*, vol. 66, no. 11, pp. 2085–2091, 2013.
- [17] J. Pfeiffer, M. Meißner, J. Prosiel, and T. Pfeiffer, "Classification of goal-directed search and exploratory search using mobile eye-tracking," in *Proceedings of the International Conference on Information Systems - Building a Better World through Information Systems, ICIS 2014, Auckland, New Zealand, December 14-17, 2014*, M. D. Myers and D. W. Straub, Eds. Association for Information Systems, 2014. [Online]. Available: <http://aisel.aisnet.org/icis2014/proceedings/GeneralIS/7>
- [18] F. Zhao, W. Schnotz, I. Wagner, and R. Gaschler, "Eye tracking indicators of reading approaches in text-picture comprehension," *Frontline Learning Research*, vol. 2, no. 5, pp. 46–66, 2014.
- [19] R. J. Miara, J. A. Musselman, J. A. Navarro, and B. Shneiderman, "Program indentation and comprehensibility," *Communications of the ACM*, vol. 26, no. 11, pp. 861–867, 1983.
- [20] J. Bauer, J. Siegmund, N. Peitek, J. C. Hofmeister, and S. Apel, "Indentation: simply a matter of style or support for program comprehension?" in *2019 IEEE/ACM 27th International Conference on Program Comprehension (ICPC)*. IEEE, 2019, pp. 154–164.
- [21] Google. (2021) Google java style guide. [Online]. Available: <https://google.github.io/styleguide/javaguide.html>
- [22] Microsoft. (2021) C# coding conventions. [Online]. Available: <https://docs.microsoft.com/en-us/dotnet/csharp/programming-guide/inside-a-program/coding-conventions>
- [23] I. McChesney and R. Bond, "Gaze behaviour in computer programmers with dyslexia: considerations regarding code style, layout and crowding," in *Proceedings of the Workshop on Eye Movements in Programming*, 2018, pp. 1–5.
- [24] I. Cetin, "Students' understanding of loops and nested loops in computer programming: An apos theory perspective," *Canadian Journal of Science, Mathematics and Technology Education*, vol. 15, no. 2, pp. 155–170, 2015.
- [25] R. K. Wagner, F. A. Zirps, A. A. Edwards, S. G. Wood, R. E. Joyner, B. J. Becker, G. Liu, and B. Beal, "The prevalence of dyslexia: a new approach to its estimation," *Journal of Learning Disabilities*, vol. 53, no. 5, pp. 354–365, 2020.